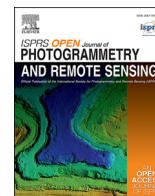Contents lists available at ScienceDirect

ISPRS Open Journal of Photogrammetry and Remote Sensing

journal homepage: www.editorialmanager.com/OPHOTO

Deep learning approach for Sentinel-1 surface water mapping leveraging Google Earth Engine

Timothy Mayer [a,b,*], Ate Poortinga [d,e], Biplov Bhandari [b,c], Andrea P. Nicolau [a,b], Kel Markert [a,b], Nyein Soe Thwal [e,f], Amanda Markert [a,b], Arjen Haag [e,g], John Kilbride [h], Farrukh Chishtie [d,e], Amit Wadhwa [i], Nicholas Clinton [j], David Saah [d,e,k]

[a] *Earth System Science Center, The University of Alabama in Huntsville, 320 Sparkman Dr., Huntsville, AL, 35805, USA*
[b] *SERVIR Science Coordination Office, NASA Marshall Space Flight Center, 320 Sparkman Dr., Huntsville, AL, 35805, USA*
[c] *Department of Atmospheric and Earth Science, The University of Alabama in Huntsville, 320 Sparkman Dr., Huntsville, AL, 35805, USA*
[d] *Spatial Informatics Group, LLC, 2529 Yolanda Ct., Pleasanton, CA, 94566, USA*
[e] *SERVIR-Mekong, SM Tower, 24th Floor, 979/69 Paholyothin Road, Samsen Nai Phayathai, Bangkok, 10400, Thailand*
[f] *Asian Disaster Preparedness Center, SM Tower, 24th Floor, 979/69 Paholyothin Road, Samsen Nai Phayathai, Bangkok, 10400, Thailand*
[g] *Deltares, Boussinesqweg 1, 2629, HV Delft, the Netherlands*
[h] *Oregon State University, USA*
[i] *World Food Programme, via Cesare Giulio Viola 68-70, 00148, Rome, Italy*
[j] *Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA, 94043, USA*
[k] *Geospatial Analysis Lab, University of San Francisco, 2130 Fulton St., San Francisco, CA, 94117, USA*

ARTICLE INFO

ABSTRACT

Satellite remote sensing plays an important role in mapping the location and extent of surface water. A variety of approaches are available for mapping surface water, but deep learning approaches are not commonplace as they are 'data hungry' and require large amounts of computational resources. However, with the availability of various satellite sensors and rapid development in cloud computing, the remote sensing scientific community is adapting modern deep learning approaches. The new integration of cloud-based Google AI platform and Google Earth Engine enables users to deploy calculations at scale. In this paper, we investigate two methods of automatic data labeling: 1. the Joint Research Centre (JRC) surface water maps; 2. an Edge-Otsu dynamic threshold approach. We deployed a U-Net convolutional neural network to map surface water from Sentinel-1 Synthetic Aperture Radar (SAR) data and tested the model performance using different hyperparameter tuning combinations to identify the optimal learning rate and loss function. The performance was then evaluated using an independent validation data set. We tested 12 models overall and found that the models utilizing the JRC data labels showed a better model performance, with F1-scores ranging from 0.972 to 0.986 for the training test and validation efforts. Additionally, an independently sampled high-resolution data set was used to further evaluate model performance. From this independent validation effort we observed models leveraging JRC data labels produced F1-Scores ranging from 0.9130.922. A pairwise comparison of models, through varying input data, learning rates, and loss functions constituents, revealed the JRC Adjusted Binary Cross Entropy Dice model to be statistically different than the 66 other model combinations and displayed the highest relative evaluations metrics including accuracy, precision score, Cohen Kappa coefficient, and F1-score. These results are in the same range as many of the conventional methods. We observed that the integration of Google AI Platform into Google Earth Engine can be a powerful tool to deploy deep-learning algorithms at scale and that automatic data labeling can be an effective strategy in the development of deep-learning models, however independent data validation remains an important step in model evaluation.

1. Introduction

Surface water is an important natural resource that sustains human wellbeing by its many purposes including drinking water, sanitation, and irrigation (Poortinga et al., 2017). Surface water is also an important component in the hydrological cycle and serves functions including electricity production, navigation, and use for industrial processes (Aekakkararungroj et al., 2020). Furthermore, it plays an important role in dictating the climate, biological diversity, and land conservation practices (Tockner and Stanford, 2002; Kong et al., 2017; Valentin et al., 2008). However, the occurrence of surface water also has negative connotations, for example in the habitat characteristics and resulting occurrence of vector borne diseases as well as disasters such as floods and drought (Dom, 2019). As such, it is evident that insight to the location and extent of surface water is critical in the context of sustainable water management. Additionally, the capability to rapidly assess, map, and disseminate impacted areas is essential to assisting national and local governments, NGOs and emergency services, which enables information to be gathered over large distances and visualized for both disaster preparedness and planned response efforts (Nemni et al., 2020; Phongsapan et al., 2019).

Satellite remote sensing has traditionally been used to map the location and extent of water surfaces. There are numerous approaches including spectral indices (Gao, 1996), machine learning technologies (Huang et al., 2018), and dynamic thresholding (Tiwari et al., 2020; Markert et al., 2020). Furthermore, a variety of active and passive sensors have been used to study surface water occurrence. Whereas passive sensors rely mostly on the visible and infrared part of the electromagnetic spectrum, active sensors use the microwave spectrum (Flores-Anderson et al., 2019). A notable effort to map surface water was done by Pekel et al. (2016). They created a planetary scale surface water time-series using the Landsat legacy data-series for the past 3 decades. However, these maps have a medium-spatial resolution and data is impeded by atmospheric conditions. This is a major issue in tropical areas with persistent cloud cover. More recently launched satellites use active space-borne microwave remote sensing. Data from these satellites have a finer resolution and are not affected by cloud cover (Oddo and Bolten, 2019). The use of UAVs has dramatically increased due to their relatively low-cost and high-operational capability to rapidly capture images and generate high resolution map products (Bhandari et al., 2015; Osco et al., 2021). The application of UAVs alongside remotely sensed data is on the rise (Emilien et al., 2021; Easterday et al., 2019), in particular the use of UAVS for water extraction mapping is a growing field [20,?, 21], however the resources and costs associated with sensor calibration and image assemblage are a frequent challenge.

Deep learning and big data analytics have become commonplace in many scientific disciplines. This paradigm shift is also quickly evolving in the field of satellite remote sensing. However, deep learning techniques are notorious for being 'data hungry' and have large computational demands (Miko lajczyk and Grochowski, 2018; Kaushal et al., 2019). In the context of applied Earth observations, there is a growing wealth of data with added location, time, and multi-modal data (e.g. active and optical) components (Zhu et al., 2017). Cloud-based geo-computational platforms such as Google Earth Engine (GEE) have resolved many of the data management and computational challenges by centralizing and standardizing data into a common framework reducing barrier to use Earth Observation data (Gorelick et al., 2017). GEE has been leveraged in numerous scientific studies (Tassi and Vizzari, 2020; Campos-Taberner et al., 2018; Aguilar et al., 2018; Parks et al., 2018) and is also used for operational purposes (Uddin et al., 2019; Markert et al., 2018; Poortinga et al., 2018). The recent integration of big data with deep learning technologies enables utilization by a wide variety of users including those across scientific disciplines.

Methods and terminologies for reference data labeling (also referred to as data collection), training models, and image classification (also referred to as inference) are disciplinary and difficult to adapt to other fields. Currently there is a large variety of model repositories with pre-trained models and large hierarchical databases (Deng et al., 2009) with associated labeled data, however, these mostly include common-place objects and are built around the RGB channels of conventional cameras. Moreover, reference data collection campaigns have traditionally used point data for image classification schemes (Saah et al., 2019a). Deep learning approaches can leverage image patches, which we defined as a 256×256 neighborhood, for image segmentation and object detection algorithms (Sharma et al., 2017). Additionally, Earth observation satellites operate in different parts of the electromagnetic spectrum and are not constrained to visible light. Remote sensing analyses are further complicated by noise measurement caused by ephemeral variations in atmospheric conditions, sensor characteristics, or background sources that can negatively affect the performance of classification algorithms. Efficiently collecting large amounts of training image patches could significantly speed up the development of neural networks. However, combining different data sources presents a challenge in terms of spatio-temporal alignment and consistency.

Although the integration of GEE with Google AI platform enables users to deploy deep learning technologies and approaches at unprecedented scales, it remains a challenge to develop these models due to the significant data requirements, the computational cost, and the degrees of freedom in the model. Moreover, deep learning methodologies often split the data in three components of training, testing, and validation, where the latter is reported as an independent measure for accuracy. In this study we use the GEE computational platform to map surface water from Sentinel-1 Synthetic Aperture Radar (SAR) data (Torres et al., 2012). The study has three main objectives: (1) to study and compare two methods of automatic data labeling for training, testing, and validating deep learning models; (2) to perform an extensive hyperparameter comparison to identify the optimal learning rate and loss function; (3) to conduct an independent validation leveraging higher resolution data to compare with the reported model results. The current study should help guide the remote sensing community in developing robust strategies for data labeling, model development, and model validation.

2. Methods

2.1. Study area

The study was conducted in Cambodia (Fig. 1), a country located in Southeast Asia with a population of approximately 16 million people (CIESIN, 2016). Cambodia has a tropical monsoon climate with most rainfall occurring between the months June and September (Misra and DiNapoli, 2014). Cambodia is located in the lower downstream part of the Mekong river basin. The Mekong river is the lifeblood for a large portion of the Khmer people who heavily rely on agriculture for their livelihoods. Moreover, Cambodia is host to the Tonle Sap, the largest fresh water lake in Southeast Asia. The Tonle Sap lake has an extremely productive ecosystem, but also serves as a flood buffer for the lower Mekong basin (Kummu et al., 2014). Monitoring and understanding surface water dynamics is important for flood disaster response but also for the protection of a valuable ecosystem. The construction of upstream reservoirs is likely to impact natural flow dynamics (Aekakkararungroj et al., 2020). The Lower Mekong region experiences a high percentage (>50%) annual mean cloud frequency, with a relatively low annual cloud variability as well (Wilson and Jetz, 2016). SAR imagery enables effective mapping and monitoring of surface water dynamics on a regular interval without the impediment of persistent cloud cover (Sanyal and Lu, 2004).

2.2. Sentinel-1 data

The Sentinel-1 satellites carry a C-band SAR sensor. This sensor can operate in multiple acquisition modes at different ground sampling distances (GSD). We utilized the Sentinel-1 Level-1 Interferometric Wide

Fig. 1. Study Area in Southeast Asia focused on Cambodia.

swath (IW) Ground Range Detected (GRD) with spatial resolution (rg x az m) at 20 × 22 and pixel spacing (rg x az m) 10 × 10 (Potin et al., 2012) specifically leveraging the dualpolarization with a vertical transmitting with vertical receiving (VV) and vertical transmitting with horizontal receiving (VH). All tiles were processed as described by (Markert et al., 2020) before ingesting them into GEE. Specifically, custom processing was done for each tile using the Sentinel-1 SNAP7 Toolbox (Sentinel Application Platform, http://step.esa.int/main/toolboxes/snap/), where the Digital Elevation Model (DEM) data from the Shuttle Radar Topography Mission (SRTM) (Farr et al., 2007) was used to perform radiometric terrain correction (RTC) and geocoding. RTC processing was based on the pixel-area integration algorithm by (Small, 2011). Additional pre-processing steps were conducted to provide the radar backscatter data set in *dB* units. Additionally, a Lee-sigma speckle.

Filter (Lee et al., 2009) was applied which is not included as standard pre-processing step in the Sentinel-1 product in the GEE data catalog.

In this study we not only included the backscatter observations from VV and VH channels but also the indices shown in Table 1. Where σ^0 is the sigma naught backscatter coefficients for VH polarization and σ^0 is the sigma naught backscatter coefficients for VV polarization. Employing SAR indices are critical because they are generated from a combination of radar measurements, which can improve the sensitivity for estimating and/or monitoring a surface characteristic such as landcover or surface water (Flores-Anderson et al., 2019; Kim et al., 2011; Huang et al., 2017).

2.3. Training data

Training data were collected for the year 2018. Two different methods for data labeling were applied. For the first method, the

Table 1

SAR Indices based on Sentinel-1 backscatter data.

Index	Definition
Polarized Ratio (VHrVV) ((Huang et al., 2018; Brisco et al., 2011))	$\dfrac{\sigma^0_{VH}}{\sigma^0_{VV}}$
Normalized Difference Polarized Index (NDPI) ((Huang et al., 2018; Mitchard et al., 2012))	$\dfrac{\sigma^0_{VV} - \sigma^0_{VH}}{\sigma^0_{VV} + \sigma^0_{VH}}$
Normalized VH Index (NVHI) ((Huang et al., 2018; McNairn and Brisco, 2004))	$\dfrac{\sigma^0_{VH}}{\sigma^0_{VV} + \sigma^0_{VH}}$
Normalized VV Index (NVVI) ((Huang et al., 2018; Charbonneau et al., 2005))	$\dfrac{\sigma^0_{VV}}{\sigma^0_{VV} + \sigma^0_{VH}}$
Radar Vegetation Index (RVI) ((Charbonneau et al., 2005; Nasirzadehdizaji et al., 2019; Yamada, 2015))	$\dfrac{4\sigma^0_{VH}}{\sigma^0_{VV} + \sigma^0_{VH}}$

Sentinel-1 data was combined with the Joint Research Centre (JRC) Monthly Water History, EC JRC/Google (Pekel et al., 2016) available in the GEE data catalog. We refer to this input data set as "JRC". The Landsat 5, 7, and 8 derived JRC data set contains the location and temporal distribution of surface water from 1984 to 2019 with water, land, and no-data as specified classes at a 30 m spatial resolution. No-data is linked to observations that contain clouds, cloud shadows, or other artifacts; no-data area were masked out and not used in calculations.

To collect training data, 100 random stratified points were distributed across the scene extent, balanced between water and land classes. This process was repeated for a total of 179 scenes observed in Fig. 2. The JRC water map of the respective month when the image was taken was included. We then buffered the individual points to construct the aforementioned image patches, to match a 256 × 256 square at a 10 m spatial
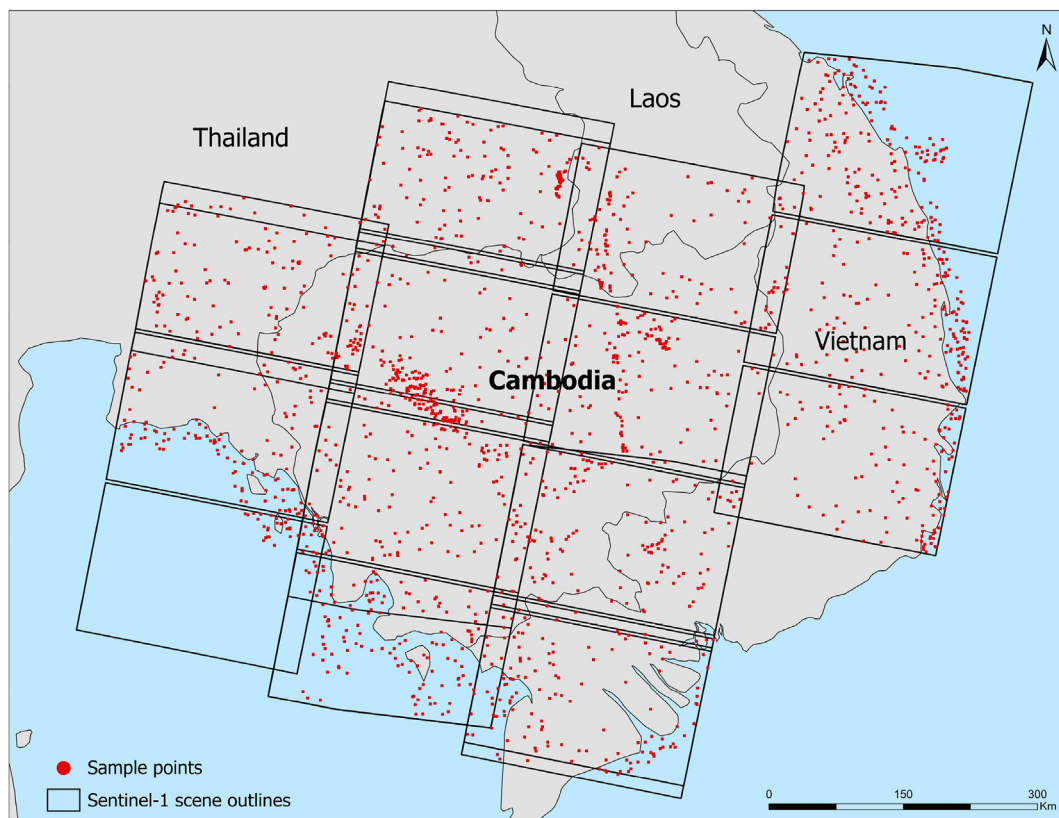
Fig. 2. Data labels were collected from the JRC data set and the Edge-Otsu method. A total of 8843 random stratified points were placed on 179 scenes, generating the 8843 patches. These were sampled on a 256 × 256 pixel window.

resolution at each point. We sampled the JRC image and removed all patches that contained no-data pixels. We used a total of 179 scenes and included a total of 8843 patches with a 256 × 256 neighborhood in the analysis.

Due to temporal and spatial inconsistencies between the Landsat derived JRC water data and Sentinel-1 SAR data, for the second data labeling approach, using the VV polarization, we applied a dynamic thresholding method to create binary land and water maps for each Sentinel-1 scene. Specifically to the Sentinel-1 SAR data set we applied the Edge-Otsu algorithm and refer to this input data set as "Edge" (Donchyts et al., 2016; Markert et al., 2020). The algorithm uses an index highlighting water to extract edges utilizing a Canny edge filter (Canny, 1986), which were then buffered, and sampled as input for Otsu thresholding (Otsu, 1979). The binary water/non-water maps were then sampled using the same collection of 256 × 256 patches, for all scenes, to ensure a consistent data series for comparison. For a complete description and application of the Edge algorithm see Markert et al., 2020 (Markert et al., 2020).

2.4. Data processing

The data processing workflow is shown in Fig. 3. The labels were combined with the relevant Sentinel-1 imagery and exported as TensorFlow (TF) records to Google Cloud Storage. These records were then imported into a virtual machine with 24 CPU cores, 224 GB memory and 4 T M60 graphic.

Cards. The hyperparameter tuning was conducted on the Graphical Processing Units (GPU). The set of 12 derived models were then exported to the Google AI platform. The *ee.Model.fromApiPlatformPredictor* function in GEE was used to import the model and conduct the inference. The integration between Google AI Platform and GEE enables large data processing, however, there is financial cost associated with the inference

and training.

2.5. Model architecture

The model used in this analysis, depicted in Fig. 4, was inspired by the U-Net architecture (Ronneberger et al., 2015). The encoder component of the model is adapted from the Visual Geometry Group (VGG) 19 model architecture (Simonyan and Zisserman, 2014). This model architecture is comprised of five multiple convolution layer encoding blocks with a distinct max pooling layer at the end of each block. This configuration ultimately increases the feature space and reduces the image resolution. For the decoder component of the architecture, a custom decoder was developed which consists of five blocks utilizing bilinear upsampling layers, followed by convolution layers, and finally regularization layers. Recent stud-

ies found that transpose convolution layers for upsampling efforts produces artifacts in the network results, and by using a resize followed by convolution strategy for upsampling, results are improved (Odena et al., 2016; Wojna et al., 2019). Each convolution layer in the decoder was initialized using a He Normal initialization (He et al., 2015) and is followed by a Batch Normalization layer (Ioffe and Szegedy, 2015) and Rectified Linear Unit (ReLU) activation function (Nair and Hinton, 2010). The image is upsampled to the input resolution at the end of the decoder. The skip connections introduced in (Ronneberger et al., 2015), which concatenate feature maps from each encoder block to the upsampled feature maps at the beginning of each decoder block with the same spatial resolution, were included in our network architecture. This process utilizes a final exit branch consisting of a 2D spatial dropout (Tompson et al., 2015) and a final 1 × 1 convolution employing a softmax activation function.

Several regularization techniques were applied in the decoder to reduce overfitting. L2 regularization, with a rate of 1e-3, was applied to
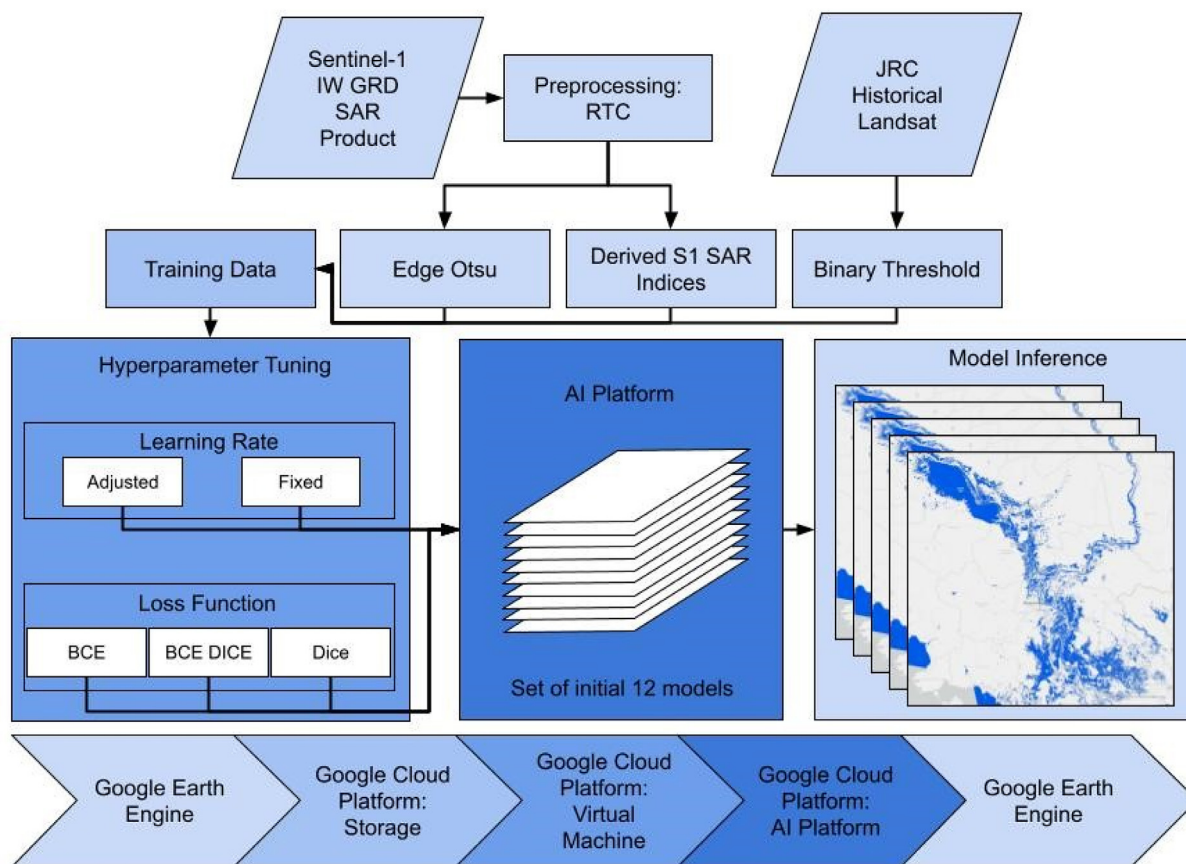
Fig. 3. Processing workflow for the generation of the 12 initial model sets. Color gradients signify the location of the processing workflow across the various platforms. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
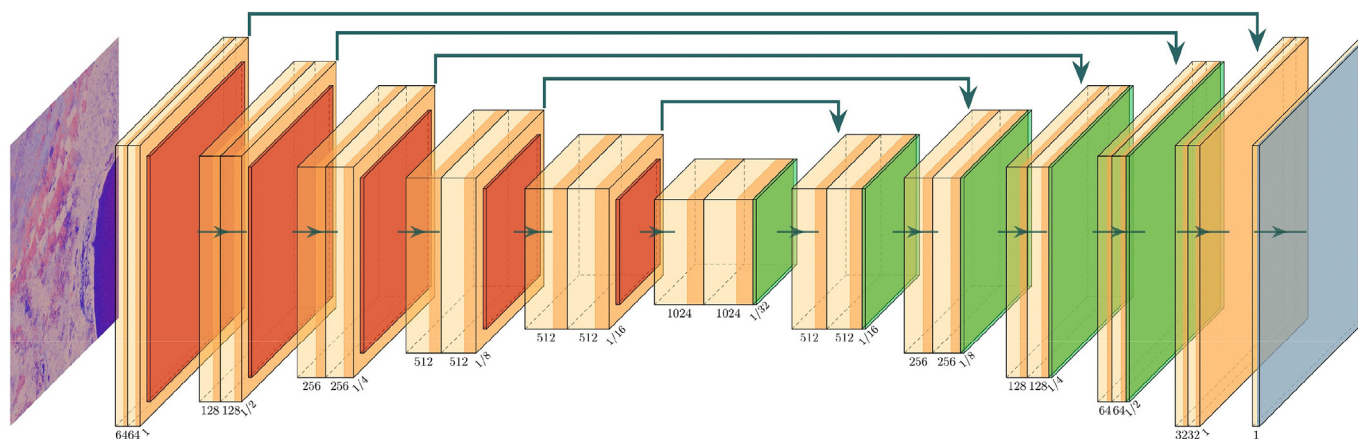

Fig. 4. The VGG19 U-Net model architecture used to map surface water. The network consists of 3 × 3 convolution layers (light orange), activation layers (dark orange), max pooling layers (red), 2D up-sampling layers (green), and an output layer (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the parameters of each convolution layer. Gaussian noise was added to the decoder block.

to reduce the over-fitting, speed up convergence, and to increase generalization (An, 1996). Finally, a dropout layer (Srivastava et al., 2014) was included after the first convolution within each decoder block. Dropout prevents the co-adaptation of neurons which yield relationships that fail to generalize outside of the training set (Srivastava et al., 2014).

Deep learning models are trained by iteratively minimizing a differential loss function. Loss functions quantify the error of predictions produced by the network with a single, scalar value. The choice of loss function can have a large impact on network performance. When conducting network training, we trained for a maximum of fifty epochs, however, we implemented an early stopping when the loss function did not improve for seven epochs to prevent overfitting. As neural networks are essentially approximations of complex functions (Liang and Srikant, 2016), the motivation for early stopping is due to a model's tendency to learn progressively more complex functions as the number of iterations increases. By limiting the time spent training the model, the complexity
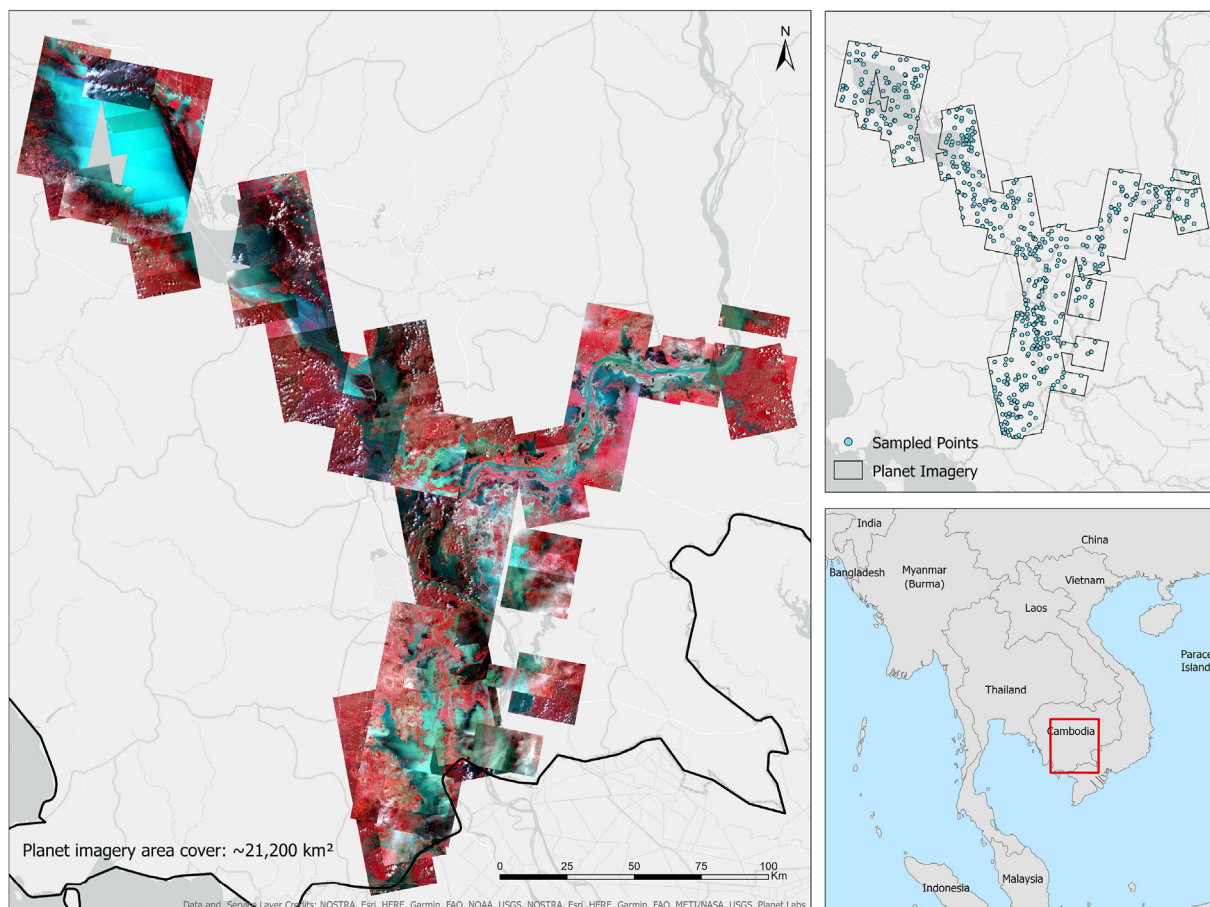
Fig. 5. Planet Scope imagery coverage with 565 sampled points. Planet images shown in false color R: NIR, G: Red, B: Green. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

of the model can be controlled, improving generalization (Yao et al., 2007). Model optimization was conducted using the Adam optimizer (Kingma and BaAdam, 2014).

For both the Edge and JRC derived training data sets, we experimented with three loss functions and two different learning rates titled as "Fixed" and "Adjusted" hereafter. For the three loss functions, we used the binary cross entropy (BCE) (Zhu et al., 2018), the dice loss function (Dice) (Milletari et al., 2016) and a combined BCE Dice loss function. For one set of loss functions we applied a Fixed learning rate (0.0001), whereas for the other set we applied an Adjusted learning rate of 0.001 for iterations below 20 epochs, a learning rate 0.0003 for iterations within 20–35 epochs, and a learning rate of 0.0001 for iterations above 35. The learning rate is an important hyperparameter. If the learning rate is too low, the network will converge slowly and will be unable to escape local minima in the loss surface. If the learning rate is too large, the network will be unable to explore minima in the loss landscape. These rates where determined through preliminary testing. This resulted in a set of 12 models. The output from these models provided a probabilistic confidence layer for the water and non-water classes. The inference was performed on a set of 11 Sentinel-1 images spanning the wet season month-month in 2019.

2.6. Accuracy assessment

The F1-score (eq (1)) was used as the metric for model performance in this study. The F1-score is calculated using the precision and recall. The precision (eq (3)) is the ratio of correctly predicted positive observations to the total predicted positive observations. The recall (eq. (2)), also referred to as sensitivity, represents the ratio of correctly predicted

positive observations to the all the observations in the class. The F1-score (eq. (1)) is the harmonic mean of precision and recall. It takes into account both the false positives and false negatives.

$$F1 = \frac{2*(Recall*Precision)}{(Recall + Precision)} \qquad (1)$$

$$recall = \frac{TP}{TP + FN} \qquad (2)$$

$$precision = \frac{TP}{TP + FP} \qquad (3)$$

where:

TP is the True Positives, which means that the actual class and the predicted class are both positive. TN is the True Negatives: which means that the actual and predicted class are both negative.

FP is the False Positives, which means that the actual class is negative whereas the predicted class is positive.

FN is the False Negative, which means that the actual class is positive but the predicted class is negative.

2.7. Performance assessment

An independent validation effort was conducted leveraging high-resolution (approximately 3 m GSD) Planet Scope visible-near infrared optical data. A total of 172 Planet Scope images within Cambodia covering an area of roughly 21,200 km^2 were utilized in the analysis. Each Planet Scope image had a corresponding Sentinel-1 SAR acquisition for the same date and area of interest. Specifically, the independent

validation data set was generated utilizing a simple random approach where sample points were distributed over the Planet Scope imagery associated with individual 2019 flood events to match the same period of the models outputs. An individual sampler performed a visual interpretation sampling approach (Lister et al., 2014; Woodward et al., 2018) on the Planet Scope imagery to estimate the presence/absence of cloud, water, and nonwater classes. The interpreter utilized a decision tree approach for classifying the validation samples (Markert et al., 2020) and the survey was constrained to a 3 by 3 pixel neighborhood to match the approximate resolution of the surface water products generated (10 m GSD). Points classified as clouds were removed leaving 565 sample points available for the validation provided in Table 2. The validation samples were then used to extract values of water/non-water from the generated surface water maps.

Additionally, for each model permutation we used a stratified K-Fold method to partition sub-samples from the larger validation data set iteratively (Kohaviet al., 1995).

The K-Fold approach allows for cross validation to estimate errors inherent for all model sets while retaining the original data set's distribution of water/non-water samples. From the generated 10 sub-samples, we calculated the Cochran's Q (Cochran, 1950) statistic for all 12 initial models, followed by a pairwise McNemar's test (McNemar, 1947) for each model constituent. Finally we calculated summary performance metrics including overall accuracy, precision, Cohen's Kappa coefficient (Cohen, 1960), and F1-score (Van Rijsbergen, 1979; Chicco and Jurman, 2020) while leveraging the independent validation data set.

3. Results and discussion

3.1. Initial model results

Utilizing the two data labeling approaches, JRC and Edge, distinct data sets were leveraged via the VGG model architecture to produce a set of 12 initial models. Table 3 shows the performance for training, testing, and validation using the BCE, Dice, and BCE Dice loss function with a Fixed and Adjusted learning rate for both input data sets, JRC and Edge. Specifically, when comparing F1-scores for the validation effort for the JRC derived models the values ranged between 0.972 and 0.975 while Edge validation ranged from 0.930 to 0.949. The models utilizing the JRC derived data set out performed all models employing the Edge data sets for the training, testing, and validation efforts as well as across all learning rates and loss functions head to head. This highlights that the JRC data labeling approach can achieve consistently higher accuracies, however, it should be noted that the Edge data labeling produced relatively high accuracies constant with Markert et al. (2020).

3.2. Independent validation results

The set of 12 initial models were further evaluated with an independent data set derived from a visual interpretation sampling approach (see Fig. 5). Specifically, the output of the initial set of models utilized a softmax activation function resulting in the surface water probability calculated for each pixel. The independent validation data set was then used to sample the probability for water and non-water. Fig. 6 shows the

Table 2

Sample points and Planet imagery distribution per date.

| Date | Sample Points | | | Planet scenes |
	Water	Not Water	Total	Total
2019–09–09	12	14	26	10
2019–09–11	44	75	119	35
2019–10–03	63	80	143	43
2019–10–05	47	92	139	51
2019–10–15	80	58	138	33
Total	246	319	565	172

Table 3

For the initial set of 12 models, F1-scores describing model performance for the variable and fixed learning rates using Dice, BCE, and BCE Dice as hyper paramter functions.

| F1-Scores | | Adjusted | | | Fixed | | |
		Dice	BCE	BCE Dice	Dice	BCE	BCE Dice
Epochs		17	16	44	9	22	18
JRC	Training	0.984	0.985	0.986	0.983	0.986	0.985
	Testing	0.975	0.975	0.976	0.973	0.975	0.975
	Validation	0.974	0.974	0.975	0.972	0.974	0.974
Epochs		29	19	12	16	22	21
Edge	Training	0.920	0.917	0.916	0.917	0.917	0.916
	Testing	0.883	0.883	0.879	0.879	0.882	0.883
	Validation	0.944	0.949	0.938	0.930	0.949	0.947

probability distributions of the 12 initial models in accordance with the independent validation data set.

We observed the median probability value for models employing JRC data sets to be higher for water classified samples across the learning rates and loss functions ranging from 0.82 to 1.00. Additionally, the interquartile range and whiskers for the JRC derived water classified samples were far more compact suggesting a heavily concentrated and sharply left skewed data prediction. However, for all 12 models we observed a dramatic left skewed probability distribution for classified water samples. For non-water classified samples, both JRC and Edge derived models displayed tight distributions with centered median values ranging between 0.00 and 0.17. We did not observe a clear visual difference between models with Adjusted or Fixed learning rates. However comparing loss functions for non-water classifications the Dice approach clearly had the tightest distribution centered at 0.00. This held true across both JRC and Edge derived models when classifying nonwater suggesting Dice as a preferable loss function for future investigations. However both BCE and BCE Dice displayed relatively strong results.

All 12 models' probability distributions displayed a large proportion of outliers evident between both classified water and non-water samples. This suggests potential inability to effectively discriminate between classes. This concern lead the research team to investigate model accuracy performance at various binary thresholds which included minimum, maximum, Q1, Q3, and 0.5 for both water and non-water classifications. We observed 0.5 as the best performing binary threshold and employed that for the remaining independent validation effort. The visually interpreted data set was randomly split 10 times into training and validation folds. The resulting analyzed splits were then averaged and the mean was used to further evaluate model performance for the set of 12 initial models displayed in Table 4.

When comparing accuracies for JRC and Edge derived models the values ranged from 0.927 to 0.936 and 0.913–0.927 respectively. JRC models' precision was consistent high across all learning rates and loss functions at 0.9770.982, while Edge models ranged from 0.990 to 1.000. JRC derived models again displayed higher Cohen Kappa Coefficient metrics 0.850–0.869 compared to Edge at 0.820–0.850. Lastly the JRC derived models offered higher F1-Scores ranging from 0.911 to 0.922, while Edge models ranged from 0.899 to 0.910. We observed a single model JRC Adjusted BCE Dice which displayed the highest accuracy, Cohen Kappa Coefficient, and F1-Score as well as relatively high precision score at 0.982.

Due to the overall high performance observed in Table 4, our team was interested in statistically identifying the best performing model from the initial set of 12 models. A Cochran Q test was performed to comparing all 12 models with the independent validation data set to determine overall model significance. We observed a p-value of 0.0003 displaying that all models did not perform equally well. From that, our team then utilized a McNemar's test in a pairwise comparison of all potential model combinations to elucidate the significant differences, results are displayed in table A5. From the statistical comparison of the 66 model combinations, we observed 9 McNemar comparisons with p-values below
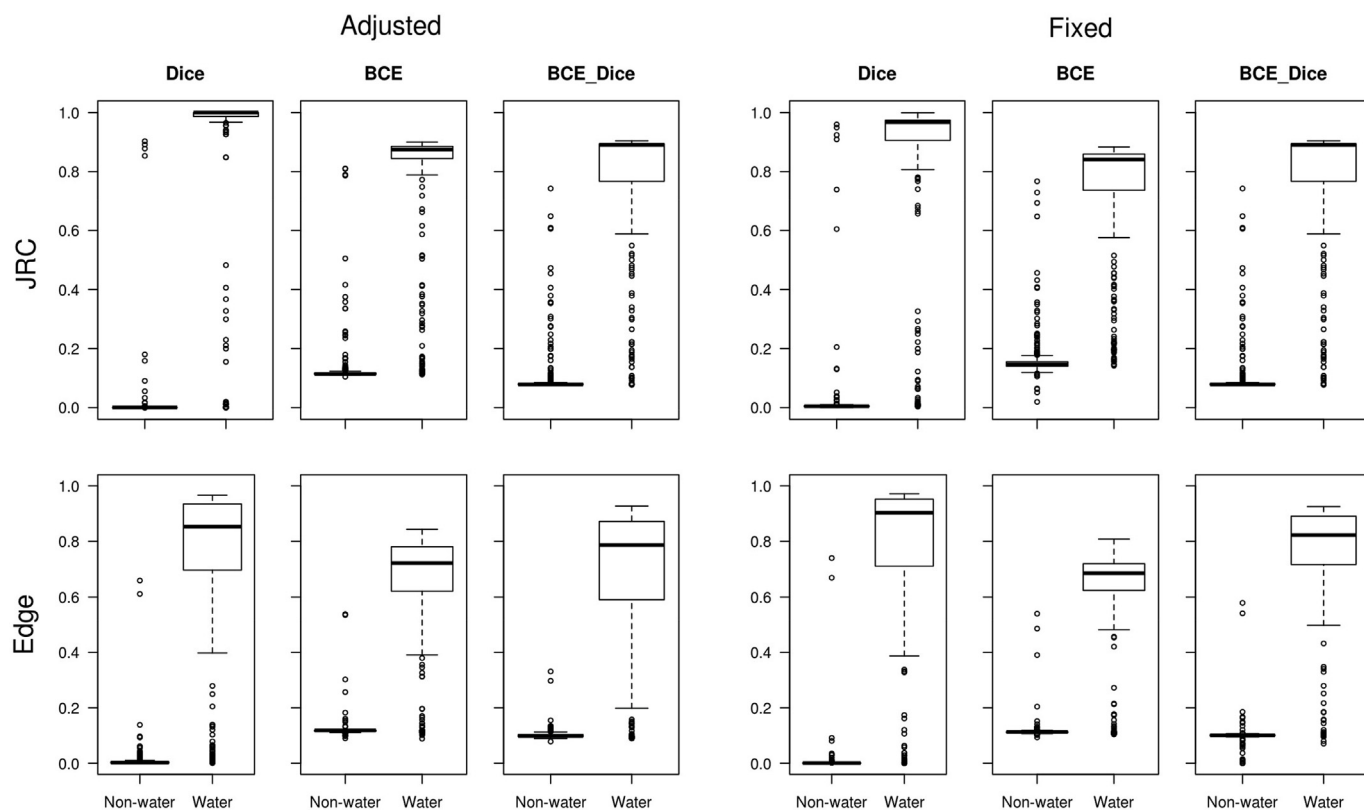
Fig. 6. Probability distribution for the 12 different models. The data was sampled from the independent validation data.

Table 4

Reported evaluation metrics from the K-folded validation split utilized in the Planet Scope imagery independent validation effort. Highlighted grey fields displayed best performing metrics.

		Adjusted			Fixed		
Validation Split		Dice	BCE	BCE Dice	Dice	BCE	BCE Dice
JRC	Accuracy	0.929	0.933	0.936	0.927	0.931	0.935
	Precision Score	0.981	0.977	0.982	0.977	0.981	0.977
	Cohen Kappa	0.854	0.861	0.869	0.850	0.858	0.865
	F1-score	0.913	0.918	0.922	0.911	0.915	0.920
Edge	Accuracy	0.927	0.927	0.913	0.920	0.924	0.927
	Precision Score	0.990	0.990	1.000	0.990	0.995	0.990
	Cohen Kappa	0.850	0.850	0.820	0.835	0.842	0.850
	F1-score	0.910	0.910	0.899	0.900	0.905	0.910

0.05, suggesting these models did not perform equally well in the independent validation effort. Observed in Fig. 7 the McNemar tables display the predictive accuracy values when utilizing the 565 available samples of the independent validation data set. Specifically, the grey upper left quadrant is correctly classified, and the grey bottom right quadrant is incorrectly classified when comparing paired models. Additionally we observed the Edge Adjusted BCE Dice model consisted of 8 of the 9 statically different combination displayed in Fig. 7 and displayed the lowest relative accuracy displayed in Table 4.

From this series of evaluations, it was identified that all of the models do display very high metrics, however, they did not perform statistically equal. This was most evident with Edge Adjusted BCE Dice, which offered the lowest metrics and was statically the most different from JRC Adjusted BCE Dice which displayed the highest overall metrics. Fig. 8 displays the JRC Adjusted BCE Dice model output for each of the associated independent validation dates, observed in Table 2, and a composite model output overlaid SAR imagery.

3.3. Caveats and limitations

While this study aims to provide a robust analysis comparing two automated data labeling approaches, with various hyperparameters, for improved operational surface water detection there are some caveats to note. First, due to the cost of cloud computing further model comparisons, including testing performance relative to classical machine learning approaches such as Random Forest were limited. Second, this study was limited due to its geographic scope and temporal range. For the independent validation effort only 565 samples, across five surveys dates were utilized. This was predominantly due to the availability of cloud-free Planet Scope imagery that coincided with Sentinel-1 imagery during the monsoon. This limited the temporal range of the validation effort. Additionally these independent validation samples were concentrated within Cambodia, while the initial set of models leveraged region wide input data. This study incorporated independent validation sample points from both riverine and lake hydrologic systems, but was ultimately limited in its investigation of diverse landscapes. SAR based mapping has been effective in flat terrain, due to the limited radiometric and geometric distortion (Horritt et al., 2003; Wickel et al., 2001). However, mountainous area with high relief subsequently cause serve radiometric distortions, and substantial correction efforts are need to reduce these errors (Song et al., 2007). Additionally, complex.

3.4. Future work

As mentioned in Section 3.3 there are additional efforts that can build upon this study. Further comparisons with classical machine learning approaches to assess both performance and ease of implementation. Performing.

an intensive ablation assessment to investigate the contribution of training data and SAR indices to model outputs. Continued evaluation of
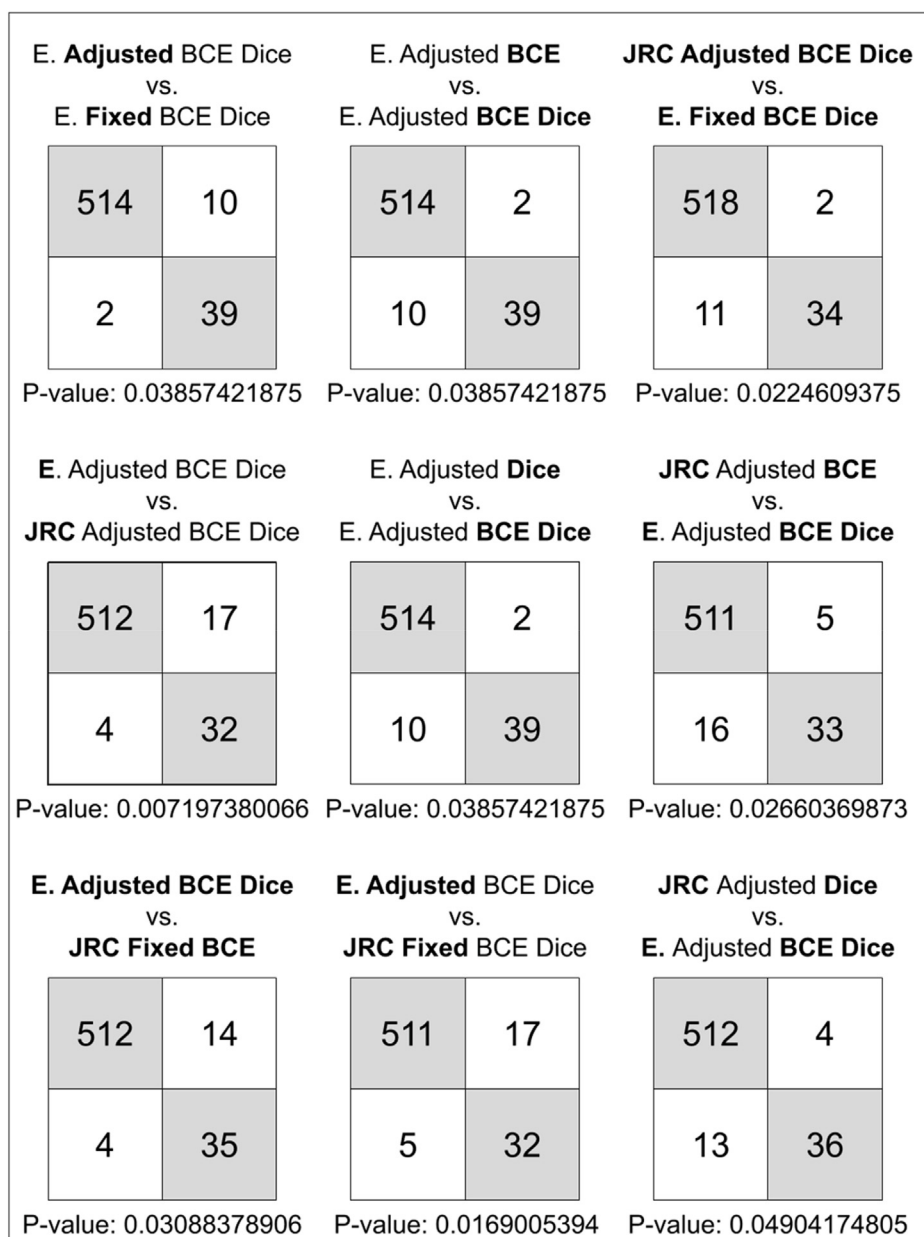
Fig. 7. From the 66 model combination, 9 model comparisons displayed p-values below 0.05, utilizing a McNemar test. These 9 significant comparisons are displayed as McNemar tables. The bold text indicates the model constituent that differs in the direct comparisons. forest and vegetation structure impact the performance of SAR based detection and classification approaches (Chapman et al., 2015; Shupe and Marsh, 2004; O'Shea et al., 2020).

the JRC data labeling approach will be explored in other regions across the globe that experience regular flooding. Employing the JRC Adjusted BCE Dice model to conduct similar thorough validation efforts through a comparison between Southeast Asia and other regions will provide crucial performance information. Further testing and utilization of this deep learning approach will be essential to integrating this workflow into an automated surface water mapping systems that provide near real-time inundation maps.

This analysis was conducted with cooperation from the SERVIR-Mekong project. SERVIR harnesses satellite and geospatial technologies to assist endusers to more effectively integrate geospatial information into their decisionmaking process. The ability to supply region wide surface water maps will further strengthen land cover monitoring (Poortinga et al., 2019a; Potapov et al., 2019; Saah et al., 2019b), food security (Poortinga et al., 2019b), and water resource management

efforts (Simons et al., 2017) in Southeast Asia.

4. Conclusion

This study explored two different data labeling methods of automatic data collection, referred to as JRC and Edge, to train a U-Net. The objectives included comparing different hyperparameters such as Adjusted and Fixed learning rates and three loss functions, Dice, BCE, BCE Dice to investigate the hyperparameters contribution to model performance. Additionally, this study utilized a rigorous independent validation process to identify the best performing model for surface water detection. The results highlighted that the JRC data labeling approach produced the best performing models. In addition, the BCE Dice loss function displayed the best overall performance. Both the Adjusted and Fixed learning rates performed similarly with no clear advantage for either approach. Overall
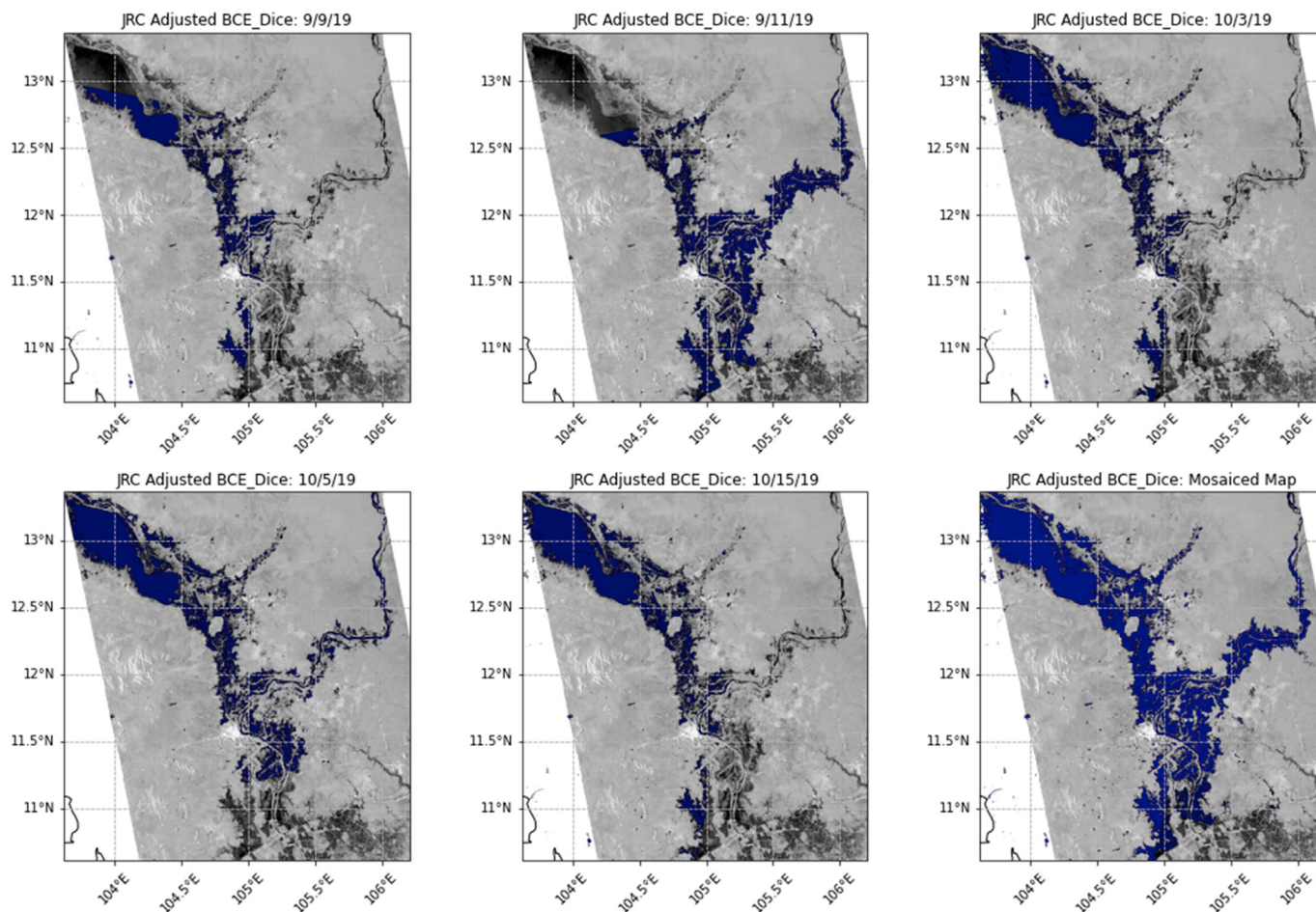
Fig. 8. The best statistically performing model, JRC Adjusted BCE Dice, visualized for each observation date and mosaiced for all dates, displays the predicted surface water.

models leveraging both JRC and Edge data sets and varied learning rates and loss functions performed well, however, a single model statistically outperformed the rest. From the pairwise McNemar comparisons we observed that the JRC Adjusted BCE Dice model provided the highest independent validation accuracy metric of 0.936 and F1-Score of 0.922. The results from this study can help inform remote sensing users on employing advanced automatic data labeling approaches by leveraging existing and freely available data sets.

The results provide insight into utilizing different hyperparameter tuning approaches and a framework for conducting an independent validation effort to more effectively identify model performance. These results contribute to improving the operational surface water map generation process. This study was conducted in collaboration and with support from the World Food Program (WFP) and Google. The WFP ingests surface water maps into their Platform for Real-time Impact and Situation Monitoring (PRISM) for flood disaster response. Through this deep learning application these results contribute to improving the rapid and automatic operational surface water mapping effort, potentially increasing the impact to beneficiaries, end-users, and stakeholders during humanitarian assistance events (Nemni et al., 2020). Additionally, integration of the Google AI platform with GEE creates a versatile technology to deploy deep learning technologies at scale. Data migration and computational demands are among the main present constraints in deploying these technologies in an operational setting. Through the implementation of these scalable technology architectures and the deep learning approaches described in this study, researchers can provide humanitarian organizations like WFP with reliable estimates of surface water throughout the monsoon season, filling a critical information gap

for humanitarian response efforts.

Author contributions

Conceptualization: T.M., A.P., K.N.M., B.B., A.N., N.T., A.M.M., A.H., F.C., A.W.; methodology: T.M., A.P., K.N.M., B.B., A.N., N.T., A.M.M.,

A. H., F.C., J.K., N.C.; data: A.P., K.N.M., B.B., A.N., N.T., T.M.,N.C., A.M.M., A.H., F.C., J.K.; software: A.P., K.N.M., B.B., A.N., N.T., N.C., T.M., A.H.; validation: T.M., A.N., B.B., A.P., N.T.; visualization, T.M., A.N., A.P., N.T., B.B.; supervision: D.S.; writing—original draft preparation, T.M., A.P., K.N.M., A.N., B.B., N.T., A.H., F.C.; writing—review & editing, T.M., A.P., B.B., A.N., K.N.M., N.T., A.M.M., A.H., J.K., F.C., A.W., N.C., D.S., and All authors have read and agreed to the published version of the manuscript.

Funding

Declaration of competing interest

The authors declare no conflict of interest. The funding agents had no

role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgments

Supplementary Materials

Source code for the processing of the raw Sentinel-1 data to RTC products is available at: https://github.com/Servir-Mekong/sentinel-1-pipeline. The source code for the Edge Otsu algorithms implemented on GEE with the.

JavaScript API is available at https://code.earthengine.google.com/?accept repo = users/kelmarkert with the code used to export this studies surface water maps available at: https://code.earthengine.google.com/eed63bb9bb36f346eeb8264c00730c7b. The source code for VGG19 U-Net model architecture with custom decoder block is available at https://github.com/Servir-Mekong/tf-vgg19-unet. An example for sampling the validation points from surface water maps and exporting the results is available at: https://code.earthengine.google.com/72e2ee06cbe46c2702f49bdc86b955b3.

Appendix A

Table A.5

McNemar pairwise comparison of all 66 models with varied input, learning rates, and loss functions. Highlighted grey fields displayed p-values below 0.05

Pairwise McNemar				JRC Adjusted		Edge Adjusted			JRC Fixed			Edge Fixed		
p-values			Dice	BCE	BCE Dic	Dice	BCE	BCE Dice	Dice	BCE	BCE Dice	Dice	BCE	BCE Dice
JRC	Adjusted	Dice	–	0.625	0.125	1.000	1.000	0.049	1.000	1.000	0.375	0.179	0.507	1.000
		BCE		–	0.625	0.507	0.507	0.026	0.25	1.000	1.000	0.092	0.266	0.507
		BCE Dice			–	0.179	0.179	0.007	0.125	0.375	1.000	0.022	0.092	0.179
Edge	Adjusted	Dice				–	1.000	0.038	1.000	0.687	0.343	0.125	0.625	1.000
		BCE					–	0.038	1.000	0.687	0.343	0.125	0.625	1.000
		BCE Dice						–	0.096	0.030	0.016	0.423	0.423	0.038
JRC	Fixed	Dice							–	0.500	0.125	0.343	0.753	1.000
		BCE								–	0.625	0.109	0.343	0.687
		BCE Dice									–	0.057	0.179	0.343
Edge	Fixed	Dice										–	0.726	0.125
		BCE											–	0.625
		BCE Dice												–

References

Aekakkararungroj, A., Chishtie, F., Poortinga, A., Mehmood, H., Anderson, E., Munroe, T., Cutter, P., Loketkawee, N., Tondapu, G., Towashiraporn, P., et al., 2020. A publicly available gis-based web platform for reservoir inundation mapping in the lower mekong region. Environ. Model. Software 123, 104552.

Aguilar, R., Zurita-Milla, R., Izquierdo-Verdiguier, E., De By, R.A., 2018. A cloud-based multi-temporal ensemble classifier to map smallholder farming systems. Rem. Sens. 10, 729.

An, G., 1996. The effects of adding noise during backpropagation training on a generalization performance. Neural Comput. 8, 643–674.

Bhandari, B., Oli, U., Pudasaini, U., Panta, N., 2015. Generation of high resolution dsm using uav images. FIG Working Week 17–21.

Brisco, B., Kapfer, M., Hirose, T., Tedford, B., Liu, J., 2011. Evaluation of c-band polarization diversity and polarimetry for wetland mapping. Can. J. Rem. Sens. 37, 82–92.

Campos-Taberner, M., Moreno-Martínez, A., García-Haro, F.J., Camps-Valls, G., Robinson, N.P., Kattge, J., Running, S.W., 2018. Global estimation of biophysical variables from google earth engine platform. Rem. Sens. 10, 1167.

Canny, J., 1986. A computation approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 8, 670–700.

Chapman, B., McDonald, K., Shimada, M., Rosenqvist, A., Schroeder, R., Hess, L., 2015. Mapping regional inundation with spaceborne l-band sar. Rem. Sens. 7, 5440–5470. https://doi.org/10.3390/rs70505440. URL: https://www.mdpi.com/2072-4292/7/5/5440.

Charbonneau, F., Trudel, M., Fernandes, R., 2005. Use of dual polarization and multi-incidence sar for soil permeability mapping. In: Advanced Synthetic Aperture Radar (ASAR). Canada, St-Hubert, QC.

Chicco, D., Jurman, G., 2020. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC Genom. 21, 1–13.

Ciesin, S., 2016. Gridded Population of the World, Version 4 (GPWV4): Population Density. Center for International Earth Science Information Network. Technical Report, Technical report.

Cochran, W.G., 1950. The comparison of percentages in matched samples. Biometrika 37, 256–266.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.

Dom, N.C., 2019. Habitat characterization of anopheles sp. mosquito larvae in malaria risk areas. Asia Pacific Environmental and Occupational Health Journal 5.

Donchyts, G., Schellekens, J., Winsemius, H., Eisemann, E., Van de Giesen, N., 2016. A 30 m resolution surface water mask including estimation of positional and thematic differences using landsat 8, srtm and openstreetmap: a case study in the murray-darling basin, Australia. Rem. Sens. 8, 386.

Easterday, K., Kislik, C., Dawson, T.E., Hogan, S., Kelly, M., 2019. Remotely sensed water limitation in vegetation: insights from an experiment with unmanned aerial vehicles (uavs). Rem. Sens. 11, 1853.

Emilien, A.-V., Thomas, C., Thomas, H., 2021. Uav & satellite synergies for optical remote sensing applications: a literature review. Science of Remote Sensing 100019.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al., 2007. The shuttle radar topography mission. Rev. Geophys. 45.

Flores-Anderson, A.I., Herndon, K.E., Thapa, R.B., Cherrington, E., 2019. The Sar Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation.

Gao, B.-C., 1996. Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. Rem. Sens. Environ. 58, 257–266.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Rem. Sens. Environ. 202, 18–27.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. ICCV.

Horritt, M., Mason, D., Cobby, D., Davenport, I., Bates, P., 2003. Waterline mapping in flooded vegetation from airborne sar imagery. Rem. Sens. Environ. 85, 271–281.

Huang, W., DeVries, B., Huang, C., Jones, J., Lang, M., Creed, I., 2017. Automated extraction of inland surface water extent from sentinel-1 data. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 2259–2262.

Huang, W., DeVries, B., Huang, C., Lang, M.W., Jones, J.W., Creed, I.F., Carroll, M.L., 2018. Automated extraction of surface water extent from sentinel-1 data. Rem. Sens. 10, 797.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift arXiv:1502.03167.

Kaushal, V., Iyer, R., Kothawade, S., Mahadev, R., Doctor, K., Ramakrishnan, G., 2019. Learning from less data: a unified data subset selection and active learning framework for computer vision. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 1289–1299.

Kim, Y., Jackson, T., Bindlish, R., Lee, H., Hong, S., 2011. Radar vegetation index for estimating the vegetation water content of rice and soybean. Geosci. Rem. Sens. Lett. IEEE 9, 564–568.

Kingma, D.P., Ba, J., Adam, 2014. A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai 14, 1137–1145. Montreal, Canada.

Kong, H., Chevalier, M., Laffaille, P., Lek, S., 2017. Spatio-temporal variation of fish taxonomic composition in a south-east asian flood-pulse system. PLoS One 12, e0174582.

Kummu, M., Tes, S., Yin, S., Adamson, P., Jozsa, J., Koponen, J., Richey, J., Sarkkula, J., 2014. Water balance analysis for the tonle sap lake–floodplain system. Hydrol. Process. 28, 1722–1733.

Lee, Jong-Sen, Wen, Jen-Hung, Ainsworth, T.L., Chen, Kun-Shan, Chen, A.J., 2009. Improved sigma filter for speckle filtering of sar imagery. IEEE Trans. Geosci. Rem. Sens. 47, 202–213. https://doi.org/10.1109/TGRS.2008.2002881.

Liang, S., Srikant, R., 2016. Why Deep Neural Networks for Function Approximation?, 04161 arXiv:1610.

Lister, T.W., Lister, A.J., Alexander, E., 2014. Land use change monitoring in Maryland using a probabilistic sample and rapid photointerpretation. Appl. Geogr. 51, 1–7.

Markert, K.N., Schmidt, C.M., Griffin, R.E., Flores, A.I., Poortinga, A., Saah, D.S., Muench, R.E., Clinton, N.E., Chishtie, F., Kityuttachai, K., et al., 2018. Historical and operational monitoring of surface sediments in the lower mekong basin using landsat and google earth engine cloud computing. Rem. Sens. 10, 909.

Markert, K.N., Markert, A.M., Mayer, T., Nauman, C., Haag, A., Poortinga, A., Bhandari, B., Thwal, N.S., Kunlamai, T., Chishtie, F., et al., 2020. Comparing sentinel-1 surface water mapping algorithms and radiometric terrain correction processing in southeast asia utilizing google earth engine. Rem. Sens. 12, 2469.

McNairn, H., Brisco, B., 2004. The application of c-band polarimetric sar for agriculture: a review. Can. J. Rem. Sens. 30, 525–542.

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12, 153–157.

Miko lajczyk, A., Grochowski, M., 2018. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW). IEEE, pp. 117–122.

Milletari, F., Navab, N., Ahmadi, S.-A., V-net, 2016. Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). IEEE, pp. 565–571.

Misra, V., DiNapoli, S., 2014. The variability of the southeast asian summer monsoon. Int. J. Climatol. 34, 893–901.

Mitchard, E.T., Saatchi, S.S., White, L., Abernethy, K., Jeffery, K.J., Lewis, S.L., Collins, M., Lefsky, M.A., Leal, M.E., Woodhouse, I.H., et al., 2012. Mapping tropical forest biomass with radar and spaceborne lidar in lop'e national park, Gabon: overcoming problems of high biomass and persistent cloud. Biogeosciences 9, 179–191.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML'10, Omnipress, Madison, WI, USA, pp. 807–814.

Nasirzadehdizaji, R., Balik Sanli, F., Abdikan, S., Cakir, Z., Sekertekin, A., Ustuner, M., 2019. Sensitivity analysis of multi-temporal sentinel-1 sar parameters to crop height and canopy coverage. Appl. Sci. 9, 655.

Nemni, E., Bullock, J., Belabbes, S., Bromley, L., 2020. Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery. Rem. Sens. 12, 2532.

Oddo, P.C., Bolten, J.D., 2019. The value of near real-time earth observations for improved flood disaster response. Frontiers in Environmental Science 7, 127.

Odena, A., Dumoulin, V., Olah, C., 2016. Deconvolution and checkerboard artifacts. Distill. URL: http://distill.pub/2016/deconv-checkerboard/.

Osco, L.P., Junior, J.M., Ramos, A.P.M., Jorge, L.A.d.C., Fatholahi, S.N., Silva, J.d.A., Matsubara, E.T., Pistori, H., Goncalves, W.N., Li, J., 2021. A Review on Deep Learning in Uav Remote Sensing. arXiv preprint arXiv:2101.10861.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 9, 62–66.

O'Shea, K., LaRoe, J., Vorster, A., Young, N., Evangelista, P., Mayer, T., Carver, D., Simonson, E., Martin, V., Radomski, P., et al., 2020. Improved remote sensing methods to detect northern wild rice (zizania palustris l.). Rem. Sens. 12, 3023.

Parks, S.A., Holsinger, L.M., Voss, M.A., Loehman, R.A., Robinson, N.P., 2018. Mean composite fire severity metrics computed with google earth engine offer improved accuracy and expanded mapping potential. Rem. Sens. 10, 879.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422. https://doi.org/10.1038/nature20584, 10.1038/nature20584.

Phongsapan, K., Chishtie, F., Poortinga, A., Bhandari, B., Meechaiya, C., Kunlamai, T., Aung, K.S., Saah, D., Anderson, E., Markert, K., Markert, A., Towashiraporn, P., 2019. Operational flood risk index mapping for disaster risk reduction using earth observations and cloud computing technologies: a case study on Myanmar. Frontiers in Environmental Science 7, 191. https://doi.org/10.3389/fenvs.2019.00191. URL: https://www.frontiersin.org/article/10.3389/fenvs.2019.00191.

Poortinga, A., Bastiaanssen, W., Simons, G., Saah, D., Senay, G., Fenn, M., Bean, B., Kadyszewski, J., 2017. A self-calibrating runoff and streamflow remote sensing model for ungauged basins using open-access earth observation data. Rem. Sens. 9, 86.

Poortinga, A., Clinton, N., Saah, D., Cutter, P., Chishtie, F., Markert, K.N., Anderson, E.R., Troy, A., Fenn, M., Tran, L.H., et al., 2018. An operational before-after-control-impact (baci) designed platform for vegetation monitoring at planetary scale. Rem. Sens. 10, 760.

Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., San Aung, K., Chishtie, F., Saah, D., 2019a. Mapping plantations in Myanmar by fusing landsat-8, sentinel-2 and sentinel-1 data along with systematic error quantification. Rem. Sens. 11, 831.

Poortinga, A., Nguyen, Q., Tenneson, K., Troy, A., Bhandari, B., Ellenburg, W.L., Aekakkararungroj, A., Ha, L.T., Pham, H., Nguyen, G.V., et al., 2019b. Linking earth observations for assessing the food security situation in vietnam: a landscape approach. Frontiers in Environmental Science 7, 186.

Potapov, P., Tyukavina, A., Turubanova, S., Talero, Y., HernandezSerna, A., Hansen, M., Saah, D., Tenneson, K., Poortinga, A., Aekakkararungroj, A., et al., 2019. Annual continuous fields of woody vegetation structure in the lower mekong region from 2000-2017 landsat time-series. Rem. Sens. Environ. 232, 111278.

Potin, P., Bargellini, P., Laur, H., Rosich, B., Schmuck, S., 2012. Sentinel-1 mission operations concept. In: 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1745–1748.

Ronneberger, O., Fischer, P., Brox, T., U-net, 2015. Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., Poortinga, A., Markert, K., Quyen, N.H., San Aung, K., et al., 2019a. Collect earth: an online tool for systematic reference data collection in land cover and use applications. Environ. Model. Software 118, 166–171.

Saah, D., Tenneson, K., Matin, M., Uddin, K., Cutter, P., Poortinga, A., Ngyuen, Q.H., Patterson, M., Johnson, G., Markert, K., et al., 2019b. Land cover mapping in data scarce environments: challenges and opportunities. Frontiers in Environmental Science 7, 150.

Sanyal, J., Lu, X., 2004. Application of remote sensing in flood management with special reference to monsoon asia: a review. Nat. Hazards 33, 283–301.

Sharma, A., Liu, X., Yang, X., Shi, D., 2017. A patch-based convolutional neural network for remote sensing image classification. Neural Network. 95, 19–28.

Shupe, S.M., Marsh, S.E., 2004. Cover-and density-based vegetation classifications of the sonoran desert using landsat tm and ers-1 sar imagery. Rem. Sens. Environ. 93, 131–149.

Simons, G., Poortinga, A., Bastiaanssen, W.G., Saah, D., Troy, D., Hunink, J., Klerk, M.d., Rutten, M., Cutter, P., Rebelo, L.-M., et al., 2017. On Spatially Distributed Hydrological Ecosystem Services: Bridging the Quantitative Information Gap Using Remote Sensing and Hydrological Models.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Largescale Image Recognition. arXiv preprint arXiv:1409, p. 1556.

Small, D., 2011. Flattening gamma: radiometric terrain correction for sar imagery. IEEE Trans. Geosci. Rem. Sens. 49, 3081–3093.

Song, Y.-S., Sohn, H.-G., Park, C.-H., 2007. Efficient water area classification using radarsat-1 sar imagery in a high relief mountainous environment. Photogramm. Eng. Rem. Sens. 73, 285–296.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

Tassi, A., Vizzari, M., 2020. Object-oriented lulc classification in google earth engine combining snic, glcm, and machine learning algorithms. Rem. Sens. 12, 3776.

Tiwari, V., Kumar, V., Matin, M.A., Thapa, A., Ellenburg, W.L., Gupta, N., Thapa, S., 2020. Flood inundation mapping-Kerala 2018; harnessing the power of sar, automatic threshold detection method and google earth engine. PLoS One 15, e0237324.

Tockner, K., Stanford, J.A., 2002. Riverine flood plains: present state and future trends. Environ. Conserv. 308–330.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C., 2015. Efficient object localization using convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 648–656.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., et al., 2012. Gmes sentinel-1 mission. Rem. Sens. Environ. 120, 9–24. https://doi.org/10.1016/j.rse.2011.05.028, 10.1016/j.rse.2011.05.028.

Uddin, K., Matin, M.A., Meyer, F.J., 2019. Operational flood mapping using multi-temporal sentinel-1 sar images: a case study from Bangladesh. Rem. Sens. 11. https://

doi.org/10.3390/rs11131581. URL: https://www.mdpi.com/2072-4292/11/13/1581.

Valentin, C., Agus, F., Alamban, R., Boosaner, A., Bricquet, J.-P., Chaplot, V., De Guzman, T., De Rouw, A., Janeau, J.-L., Orange, D., et al., 2008. Runoff and sediment losses from 27 upland catchments in southeast asia: impact of rapid land use changes and conservation practices. Agric. Ecosyst. Environ. 128, 225–238.

Van Rijsbergen, C., 1979. Information retrieval: theory and practice. In: Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems, pp. 1–14.

Wickel, A., Jackson, T., Wood, E.F., 2001. Multitemporal monitoring of soil moisture with radarsat sar during the 1997 southern great plains hydrology experiment. Int. J. Rem. Sens. 22, 1571–1583.

Wilson, A.M., Jetz, W., 2016. Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. PLoS Biol. 14, e1002415.

Wojna, Z., Ferrari, V., Guadarrama, S., Silberman, N., chieh Chen, L., Fathi, A., Uijlings, J., 2019. The Devil Is in the Decoder: Classification, Regression and Gans. IJCV. URL: https://arxiv.org/abs/1707.05847.

Woodward, B.D., Evangelista, P.H., Young, N.E., Vorster, A.G., West, A.M., Carroll, S.L., Girma, R.K., Hatcher, E.Z., Anderson, R., Vahsen, M.L., et al., 2018. Co-rip: a riparian vegetation and corridor extent dataset for Colorado river basin streams and rivers. ISPRS Int. J. Geo-Inf. 7, 397.

Yamada, Y., 2015. Preliminary study on the radar vegetation index (rvi) application to actual paddy fields by alos/palsar full-polarimetry sar data, the International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences 40, 129.

Yao, Y., Rosasco, L., Caponnetto, A., 2007. On early stopping in gradient descent learning. Constr. Approx. 26, 289–315. https://doi.org/10.1007/s00365-006-0663-2, 10.1007/s00365006-0663-2.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine 5, 8–36.

Zhu, D., Yao, H., Jiang, B., Yu, P., 2018. Negative Log Likelihood Ratio Loss for Deep Neural Network Classification. arXiv preprint arXiv:1804, p. 10690.